Poisson genericity in numeration systems with exponentially mixing probabilities

Nicolás Álvarez Verónica Becher Eda Cesaratto Martín Mereb
Yuval Peres Benjamin Weiss

August 7, 2024

Abstract

We define Poisson genericity for infinite sequences in any finite or countable alphabet with an invariant exponentially-mixing probability measure. A sequence is Poisson generic if the number of occurrences of blocks of symbols asymptotically follows a Poisson law as the block length increases. We prove that almost all sequences are Poisson generic. Our result generalizes Peres and Weiss' theorem about Poisson genericity of integral bases numeration systems. In particular, we obtain that their continued fraction expansions for almost all real numbers are Poisson generic.

1 Introduction and statement of results

Several years ago Zeev Rudnick defined the notion of Poisson genericity for real numbers: a real number is Poisson generic for an integer base b greater than or equal to 2 if in its base-b expansion the number of occurrences of blocks of digits follows the Poisson distribution. Peres and Weiss gave a metric result showing that almost all real numbers, in the sense of Lebesgue measure, are Poisson generic, see [19, 3]. A construction of a Poisson generic sequence for base 2 appears in [5].

In this note we define Poisson genericity for infinite sequences in any finite or countable alphabet with respect to an invariant probability measure that is exponentially-mixing. A sequence is Poisson generic if the number of occurrences of blocks of symbols in the sequence converges to the Poisson law. We prove that Poisson genericity holds with probability 1. Our initial goal was to prove that almost all real numbers have Poisson generic continued fractions by extending the methods developed in [19, 3]. Theorem 1 establishes this and holds in more general settings.

Let Ω be a finite or countable alphabet with at least two symbols. For each positive integer k, the set Ω^k consists of all words of length k over the alphabet Ω. We write Ω^* for the set of all finite words and $\Omega^{\mathbb{N}}$ for the set of one-sided infinite words.

We write $w = w_1 \cdots w_k$ for words in Ω^k and we use the letter x for infinite words in $\Omega^{\mathbb{N}}$. For a word w, $|w|$ is its length. We number the positions in words and infinite sequences starting from 1 and we write $w[l, r]$ for the sub-sequence of w that begins in position l and ends in position r. We use interval notation, with a square bracket when the set of integers includes the endpoint and a parenthesis to indicate that the endpoint is not included. The same convention is used for $x \in \Omega^{\mathbb{N}}$. Given a $k \in \mathbb{N}$ and a word w of length k, the subset $C(w)$ of $\Omega^{\mathbb{N}}$ defined by

$$C_k(w) = \{x \in \Omega^{\mathbb{N}} : x[1, k] = w\}$$

is called the cylinder of w. All possible cylinders of any length generate a sigma-algebra $\mathcal{B}$. Finally, we assume that a measure μ is defined on the sigma-algebra $\mathcal{B}$ so that $(\Omega^{\mathbb{N}}, \mathcal{B}, \mu)$ is a probability space. For every $k \in \mathbb{N}$, the projection of μ over the first k coordinates is a measure on Ω^k that we denote by μ_k. To shorten notation, for a word $w \in \Omega^k$ we write

$$\mu_k(w) = \mu(C_k(w))$$

and for $W \subseteq \Omega^k$ we write

$$\mu_k(W) = \sum_{w \in W} \mu_k(w).$$

For $j \in \mathbb{N}$, $x \in \Omega^{\mathbb{N}}$, $k \in \mathbb{N}$ and $w \in \Omega^k$, we write $I_j(x, w)$ for the indicator function that the word w occurs in the sequence x starting at position j,

$$I_j(x, w) = \mathbb{1}_{x[j,j+k)=w}.$$

For each $k \in \mathbb{N}$, we define on the space $\Omega^{\mathbb{N}} \times \Omega^k$ with measure $\mu \times \mu_k$ the integer-valued random measure $M_k = M_k(x, w)$ on $\mathbb{R}^+ = [0, +\infty)$ by

$$M_k(x, w)(S) = \sum_{j \colon j\mu_k(w) \in S} I_j(x, w), \quad \text{for any Borel set } S \subseteq \mathbb{R}^+.$$

We also define, for each $x \in \Omega^{\mathbb{N}}$ and $k \in \mathbb{N}$, the integer-valued random measure M_k^x as

$$M_k^x(w)(S) = M_k(x, w)(S) \quad \text{for any Borel set } S \subseteq \mathbb{R}^+.$$

Similarly, for each $k \in \mathbb{N}$ and each fixed $w \in \Omega^k$, we have the integer-random measure given by

$$M_k^w(x)(S) = M_k(x, w)(S) \quad \text{for any Borel set } S \subseteq \mathbb{R}^+.$$

A point process $Y(\cdot)$ on $\mathbb{R}^+$ is an integer-valued random measure. Therefore, $M_k(\cdot)$ and $M_k^x(\cdot)$ are point processes on $\mathbb{R}^+$ for each $k \geq 1$. The Poisson point process on $\mathbb{R}^+$ is a point process $Y(\cdot)$ on $\mathbb{R}^+$ such that the following two conditions hold: (a) for all disjoint Borel sets $S_1, \ldots, S_m$ included in $\mathbb{R}^+$, the random variables $Y(S_1), \ldots, Y(S_m)$ are mutually independent; and (b) for each bounded Borel set $S \subseteq \mathbb{R}^+$, $Y(S)$ has the distribution of a Poisson random variable with parameter equal to the Lebesgue measure of S. A sequence $(Y_k(\cdot))_{k \geq 1}$ of point processes converges in distribution to a point process $Y(\cdot)$ if, for every Borel set S, the random variables $Y_k(S)$ converge in distribution to $Y(S)$ as k goes to infinity. A thorough presentation on Poisson point processes can be read from [14] or [17].

We write $p(\lambda, j)$ to denote the Poisson mass function $e^{-\lambda}\lambda^j/j!$.

Definition (Poisson genericity)**.** *We say that $x \in \Omega^{\mathbb{N}}$ is Poisson generic if the sequence $(M_k^x(.))_{k \in \mathbb{N}}$ of point processes on $\mathbb{R}^+$ converges in distribution to a Poisson point process on $\mathbb{R}^+$, as k goes to infinity. This means that for every Borel set $S \subseteq \mathbb{R}^+$, every integer $j \geq 0$,*

$$\mu_k(\{w \in \Omega^k : M_k^x(w)(S) = j\}) \to p(|S|, j), \ \text{as } k \to \infty.$$

Our main result is Theorem 1 and it holds under the following assumptions on the measure μ:

Exponentially ψ-mixing: For each $k \in \mathbb{N}$ and $\ell \in \mathbb{N}$ such that $1 \leq k < \ell$ or $\ell = \infty$, consider the sigma-algebra $\mathcal{B}_{k,\ell}$ generated by the sets

$$\{x \in \Omega^{\mathbb{N}} : I_k(x, w) = 1\}$$

where w is any word in $\Omega^{\ell-k}$ for finite ℓ or in $\Omega^\star$ for $\ell = \infty$. The measure μ is mixing if there are constants $\sigma \in (0, 1)$ and $T > 0$ so that for every $k, \ell \in \mathbb{N}$ where $\ell > k$, for every $A \in \mathcal{B}_{1,k}, B \in \mathcal{B}_{\ell,\infty}$ of non-zero measure,

$$\left| \frac{\mu(A \cap B)}{\mu(A)\mu(B)} - 1 \right| \leq T\sigma^{\ell-k}. \tag{1}$$

In terms of words and indicator functions, this mixing property implies that for every $u, v \in \Omega^*$, for every $i, j \in \mathbb{N}$ where $j > i + |u|$,

$$\left| \frac{\mu(x \in \Omega^{\mathbb{N}} : I_i(x,u) I_j(x,v) = 1)}{\mu_{|u|}(u)\mu_{|v|}(v)} - 1 \right| \leq T\sigma^{j-i-|u|}. \tag{2}$$

Invariance: The measure μ is invariant if for every $k \in \mathbb{N}$, $w \in \Omega^k$ and $i, j \in \mathbb{N}$,

$$\mu(x \in \Omega^{\mathbb{N}} : I_i(x,w) = 1) = \mu(x \in \Omega^{\mathbb{N}} : I_j(x,w) = 1). \tag{3}$$

Theorem 1. *For any invariant and exponentially ψ-mixing probability measure μ on $\Omega^{\mathbb{N}}$, μ-almost all $x \in \Omega^{\mathbb{N}}$ are Poisson generic.*

Thus, Theorem 1 says that for μ-almost all $x \in \Omega^{\mathbb{N}}$, for every Borel set $S \subseteq \mathbb{R}^+$, for every integer $j \geq 0$,

$$\mu_k\left(\{w \in \Omega^k : M_k(x,w)(S) = j\}\right) \to p(|S|, j), \text{ as } k \to \infty.$$

This result is dual to metric results proved in the context of dynamical systems which say that for each $t > 0$, there exists an exceptional set $E_t \subset \Omega^{\mathbb{N}}$ with $\mu(E_t) = 0$ so that for any $y \in \Omega^{\mathbb{N}} \setminus E_t$ and every integer $j \geq 0$,

$$\mu\left(\left\{x \in \Omega^{\mathbb{N}} : M_k(x, y[1,k])((0,t)) = j\right\}\right) \to p(t,j), \text{as } k \to \infty.$$

The difference between the two approaches is whether the point process draws at random the first or the second argument of $M_k(x,w)$. The symmetry in the definition of the statements does not yield a symmetry of the respective proofs. Each of the two cases require their own proof.

Also, we remark that Theorem 1 proves that there is an universal exceptional set $E \subset \Omega^{\mathbb{N}}$, with $\mu(E) = 0$ such that the Poisson law holds for every $x \in \Omega^{\mathbb{N}} \setminus E$ and every Borel set $S \in \mathbb{R}^+$ (in particular, for $S = (0,t)$).

The work on limit theorems for mixing sequences dates back to Doeblin in 1940 [8] with his statement that the number of occurrences of large digits in the partial quotients of continued fractions follows the Poisson law. This was later proved by Iosifescu [11], see also [9]. These pioneering works on continued fractions have evolved into the study of the statistics of the number of visits of orbits under a given discrete dynamical system to a sequence of sets of positive measures shrinking to a point. When the sequence of visited sets is sufficiently regular and the dynamical system satisfies mixing conditions, they prove that the number of visits follows a Poisson Law as the measures of the sets converge to zero. Since the first papers by Collet, Coelho, Galves, Hirata and Schmitt, the subject has developed into many directions, for instance, the study of families of dynamical systems with different mixing conditions, the statistics of periodic orbits, the study of error terms. We refer to [10] and [20] for references on this subject with brief historical accounts.

Within dynamical systems, our Theorem 1 holds for fibred numerations systems with an invariant probability measure satisfying the mixing condition (1). Following [4, Definitions 2.3], given a compact set X and a map $T : X \to X$, we say that (X,T) is a fibred system if the transformation $T \colon X \to X$ is such that there exist a finite or countable set Ω and a topological partition $P = \{C(a)\}_{a \in \Omega}$ of X for which the restriction T_a of T to $C(a)$ is injective and continuous, for each $a \in \Omega$. (Here topological partition means that sets $C(a)$ for $a \in \Omega$ are non-empty, open, and disjoint, and the union of their closures is the whole X.) As it is proved in [4], a fibred system defines a numeration system called fibred numeration system. Examples of fibred systems with an invariant probability measure and satisfying our mixing conditions are the following: integer bases and its generalization to numeration systems induced by a finite or countable set of digits which are independent, the beta shift, the classical, centered, and odd continued fractions algorithms, and in general, strongly expanding maps of the interval.

The map associated with continued fractions is the Gauss map

$$T : [0,1] \to [0,1], \quad T(x) = 1/x - \lfloor 1/x \rfloor, \text{ if } x \neq 0 \text{ and } T(0) = 0.$$

The Gauss measure defined as $d\mu(x) = dx/((1+x)\ln 2)$ is invariant for T and is exponentially ψ-mixing (see [12]). There are also examples in two dimensions as the Ostrowski dynamical system. In [4], it is proved that it is fibred, and in[6, Theorem 4.4] it is proved that it has an invariant and exponentially ψ-mixing probability measure.

Another point view for numerations systems is to consider them as stochastic processes taking values in a finite or countable alphabet such as irreducible and aperiodic Markov chains with a finite number of states. Theorem 1, of course, applies to these cases.

To prove Theorem 1 we adapt Peres and Weiss' proof strategy used in [19, 3]. The proof consists of two parts, an annealed result and a quenched result.

For the annealed result, for each word in Ω^k we use the Chen–Stein method [7] to obtain a uniform bound for the rate of convergence of $M_k^w(.)$ to the Poisson law. We then show that the sequence of point processes $M_k(.)$ on $\mathbb{R}^+$ converges in distribution to a Poisson point process on $\mathbb{R}^+$ in the product space $\Omega^{\mathbb{N}} \times \Omega^k$. To obtain this part we prove Lemma 4, which is a version for an infinite alphabet of Abadi and Vergne's pointwise limit theorem with sharp error terms [2] which holds for a finite alphabet.

The quenched result is an application of a concentration inequality that proves that what happens on average in the space $\Omega^{\mathbb{N}}$ is essentially what happens for almost all sequences. With this end in mind we prove a concentration inequality that holds for countable alphabets and for functions depending on countably many variables with the bounded differences property known as Lipschitz condition for a weighted Hamming distance. For this we adapt the martingale difference method given in [15, Theorem 3.3.1] and [16, Theorem 1.1]. The former deals with a finite alphabet and weighted Hamming distances and the latter deals with an infinite alphabet but a constant Hamming distance.

To account for infinite memory numeration systems with digits in a countable alphabet we have to work with functions depending on infinitely many variables.

2 Proof of Theorem 1

The next proposition gives sufficient conditions for a sequence of random measures on $\mathbb{R}^+$ to converge to a Poisson random measure on $\mathbb{R}^+$. These are conditions just on integer-valued random measures (point processes) on finite unions of disjoint intervals with rational endpoints. It is later used in Lemmas 7 and 10.

Proposition 1 (Instantiation of Kallenberg [13, Theorem 4.18]). *Let $(X_k(\cdot))_{k\in\mathbb{N}}$ be a sequence of point processes on $\mathbb{R}^+$ and let $Y(\cdot)$ be a Poisson process on $\mathbb{R}^+$. If for any $S \subseteq \mathbb{R}^+$ that is a finite union of disjoint intervals with rational endpoints we have*

1. $\limsup_{k\to\infty} \mathbb{E}[X_k(S)] \le \mathbb{E}[Y(S)]$ *and*
2. $\lim_{k\to\infty} \mathbb{P}\left(X_k(S) = 0\right) = \mathbb{P}\left(Y(S) = 0\right)$

then $X_k(\cdot)$ converges in distribution to $Y(\cdot)$, as $k \to \infty$.

2.1 The annealed result

Fix a real number $\lambda > 0$. For any $h : \mathbb{N} \to \mathbb{R}$, the number $P(\lambda, h)$ denotes the average value of h with respect to the Poisson distribution of parameter λ:

$$P(\lambda, h) = e^{-\lambda} \sum_{j=0}^{\infty} h(j)\lambda^j/j!.$$

In particular, if $h(j) = \mathbb{1}_j$, we recover the probability mass function $p(\lambda, j)$, where $p(\lambda, j) = e^{-\lambda}\lambda^j/j!$. When W is a random variable defined on $\Omega^{\mathbb{N}}$, we write

$$\mathbb{E}_\mu\left[W\right] = \int_{\Omega^{\mathbb{N}}} W(x)d\mu(x).$$

If $W = \mathbb{1}_S$ for some $S \subset \Omega^{\mathbb{N}}$, $\mathbb{E}_\mu[\mathbb{1}_S] = \mu(S)$. Similarly, when W is defined on $\Omega^{\mathbb{N}} \times \Omega^k$, we write

$$\mathbb{E}_{\mu \times \mu_k}[W] = \int_{\Omega^{\mathbb{N}} \times \Omega^k} W(x, w) d\mu(x) d\mu_k(w).$$

Every mixing measure μ has a contraction ratio, [1, Lema 1].

Definition (Contraction ratio ρ)**.** *Let an invariant mixing measure μ on $\Omega^{\mathbb{N}}$. The measure μ has a contraction ratio $\rho \in (0, 1)$ if there is constant $K > 0$ such that for every $k \in \mathbb{N}$, for every $w \in \Omega^k$,*

$$\mu_k(w) \leq K\rho^k. \tag{4}$$

Definition (Bounded distortion constant R)**.** *The mixing property for μ implies that here is a constant $R > 0$ such that, for every $u, v \in \Omega^*$, for every $i, \ell \geq 1$,*

$$\mu(x \in \Omega^{\mathbb{N}} : I_i(x, w) I_{i+\ell}(x, v) = 1) \leq R\mu_{|u|}(u)\mu_{|v|}(v).$$

We call this the bounded distortion property and we say that R is the bounded distortion constant.

Definition (Periods of a word)**.** *A word $w \in \Omega^k$ has period ℓ if $\ell < k$ and $w_i = w_{i+\ell}$ for all $1 \leq i \leq k - \ell$. For $w \in \Omega^k$, the set π_w gathers the positive integers which are its periods,*

$$\pi_w = \{\ell : w \text{ has period } \ell\}.$$

Definition (Set $\mathcal{J}_{w,S}$)**.** *For a given set $S \subset \mathbb{R}$ and a fixed word $w \in \Omega^k$, we define the set*

$$\mathcal{J}_{w,S} = \{i \in \mathbb{N} : i\mu_k(w) \in S\}.$$

We use $\#A$ to denote the cardinality of a finite set A. We are interested in the cardinality of $\mathcal{J}_{w,S}$. If S is an interval (a, b), for any $a, b \in \mathbb{R}, a < b$,

$$\frac{b-a}{\mu_k(w)} - 1 \leq \#\mathcal{J}_{w,S} \leq \frac{b-a}{\mu_k(w)} + 1.$$

Therefore, if S is a finite union of m nonempty intervals,

$$\frac{|S|}{\mu_k(w)} - m \leq \#\mathcal{J}_{w,S} \leq \frac{|S|}{\mu_k(w)} + m. \tag{5}$$

That is, $\#\mathcal{J}_{w,S} = |S|/\mu_k(w) + O(1)$ as $k \to \infty$ where the hidden constant of the O term only depends on the number of intervals of S.

Lemma 1. *Let μ be an invariant and exponentially ψ-mixing measure on $\Omega^{\mathbb{N}}$ with contraction ratio ρ. Let $S \subset \mathbb{R}^+$ be a finite union of bounded intervals. For each $k \in \mathbb{N}$, for each fixed $w \in \Omega^k$, the following hold*

1. $\mathbb{E}_\mu[M_k^w(S)] = |S| + O(\rho^k)$,
2. $\mathbb{E}_\mu\left[(M_k^w(S))^2\right] = |S| + |S|^2 + O\left(k\rho^k\right) + O\left(\sum_{\ell \in \pi_w} \rho^\ell\right)$,
3. $\mathbb{V}_\mu(M_k^w(S)) = |S| + O(k\rho^k) + O\left(\sum_{\ell \in \pi_w} \rho^\ell\right)$,

   where π_w are the periods of the word $w \in \Omega^k$ and the hidden constant in the O-term only depends on S.

Proof. We write $I_i^w(x)$ to denote $I_i(x,w)$.

Point 1. As a direct consequence of (5) we have

$$\begin{aligned}\mathbb{E}_\mu\left[M_k^w(S)\right] &= \int_{\Omega^{\mathbb{N}}} (M_k(x,w)(S))\, d\mu(x) \\ &= \sum_{i\in\mathcal{J}_{w,S}} \mathbb{E}_\mu\left[I_i^w\right] \\ &= \sum_{i\in\mathcal{J}_{w,S}} \mu_k(w) \\ &= \left(\frac{|S|}{\mu_k(w)} + O(1)\right)\mu_k(w).\end{aligned}$$

The contraction ratio property $\mu_k(w) = O(\rho^k)$ yields

$$\mathbb{E}_\mu\left[M_k^w(S)\right] = |S| + O(\mu_k(w)) = |S| + O(\rho^k).$$

Point 2. The random variable $\left(M_k^w(S)\right)^2$ is a sum which involves the products $I_i^w I_j^w$, for $i,j \in \mathcal{J}_{w,S}$. We split that sum as follows:

$$\left(M_k^w(S)\right)^2 = \sum_{i\in\mathcal{J}_{w,S}} (I_i^w)^2 + \sum_{i\in\mathcal{J}_{w,S}} \sum_{\substack{j\in\mathcal{J}_{w,S}\\ 1\le|i-j|<k}} I_i^w I_j^w + \sum_{i\in\mathcal{J}_{w,S}} \sum_{\substack{j\in\mathcal{J}_{w,S}\\ |i-j|\ge k}} I_i^w I_j^w.$$

Let

$$\begin{aligned}E_1 &= \sum_{i\in\mathcal{J}_{w,S}} \mathbb{E}_\mu\left[(I_i^w)^2\right]; \\ E_2 &= \sum_{i\in\mathcal{J}_{w,S}} \sum_{\substack{j\in\mathcal{J}_{w,S}\\ 1\le|i-j|<k}} \mathbb{E}_\mu\left[I_i^w I_j^w\right]; \\ E_3 &= \sum_{i\in\mathcal{J}_{w,S}} \sum_{\substack{j\in\mathcal{J}_{w,S}\\ |i-j|\ge k}} \mathbb{E}_\mu\left[I_i^w I_j^w\right].\end{aligned}$$

Notice that $\mathbb{E}_\mu\left[(M_k^w(S))^2\right] = E_1 + E_2 + E_3$.

We prove that $E_1 = |S| + O(\rho^k)$. This is a direct consequence of the fact that $(I_i^w)^2 = I_i^w$ because I_i^w is an indicator function and the estimate already proved for $\mathbb{E}_\mu\left[M_k^w(S)\right]$:

$$E_1 = \sum_{i\in\mathcal{J}_{w,S}} \mathbb{E}_\mu\left[(I_i^w)^2\right] = \sum_{i\in\mathcal{J}_{w,S}} \mathbb{E}_\mu\left[I_i^w\right] = \mathbb{E}_\mu\left[M_k^w(S)\right] = |S| + O(\rho^k).$$

We prove that $E_2 = O\left(\sum_{\ell\in\pi_w} \rho^\ell\right)$. The first step is based on the following observation: consider $i,j\in\mathbb{N}$, $j>i$. For a fixed w, there exists $x\in\Omega^{\mathbb{N}}$ for which $I_i^w I_j^w(x) = 1$ if and only if $j-i$ is a period of w, $j-i\in\pi_w$,

$$\sum_{i\in\mathcal{J}_{w,S}} \sum_{\substack{j\in\mathcal{J}_{w,S}\\ 1\le|i-j|<k}} \mathbb{E}_\mu\left[I_i^w I_j^w\right] = 2\sum_{\ell\in\pi_w} \sum_{\substack{i\in\mathcal{J}_{w,S}\\ i+\ell\in\mathcal{J}_{w,s}}} \mathbb{E}_\mu\left[I_i^w I_{i+\ell}^w\right].$$

The bounded distortion property, the invariance of the measure and, finally, the existence of a contraction ratio imply that

$$E_2 \le 2\sum_{\ell\in\pi_w} \sum_{\substack{i\in\mathcal{J}_{w,S}\\ i+\ell\in\mathcal{J}_{w,s}}} R\mu_k(w)\mu_\ell\left(w[1\ldots\ell]\right) \le \sum_{\ell\in\pi_w} 2RK\mu_k(w)\rho^\ell \sum_{\substack{i\in\mathcal{J}_{w,S}\\ i+\ell\in\mathcal{J}_{w,s}}} 1.$$

Now, we use (5) in order to deal with $\#\mathcal{J}_{w,s}$ and obtain

$$E_2 \leq \sum_{\ell\in\pi_w} 2RK\mu_k(w)\rho^\ell \left(\frac{|S|}{\mu_k(w)} + O(k)\right) = O\left(\sum_{\ell\in\pi_w} \rho^\ell + k\mu_k(w)\rho^\ell\right)$$
$$= O\left(\sum_{\ell\in\pi_w} \rho^\ell\right),$$

which proves the estimate.

We prove that $E_3 = |S|^2 + O(k\rho^k)$. First, by the mixing property (2)and some manipulations:

$$E_3 = \sum_{i\in\mathcal{J}_{w,S}} \sum_{\substack{j\in\mathcal{J}_{w,S}\\|i-j|\geq k}} \mathbb{E}_\mu\left[I_i^w I_j^w\right] = \sum_{i\in\mathcal{J}_{w,S}} \sum_{\substack{j\in\mathcal{J}_{w,S}\\|i-j|\geq k}} \mu_k^2(w)\left(1 + O\left(\sigma^{|i-j|-k}\right)\right)$$
$$= \mu_k^2(w)\left(\left(\sum_{i\in\mathcal{J}_{w,S}} \sum_{\substack{j\in\mathcal{J}_{w,S}\\|i-j|\geq k}} 1\right) + O\left(\sum_{i\in\mathcal{J}_{w,S}} \sum_{\substack{j\in\mathcal{J}_{w,S}\\|i-j|\geq k}} \sigma^{|i-j|-k}.\right)\right)$$

Fix $i \in \mathcal{J}_{w,S}$. Notice that $\mathcal{J}_{w,S} = (\mathcal{J}_{w,S} \cap \{j : |i-j| \geq k\}) \cup (\mathcal{J}_{w,S} \cap \{j : |i-j| < k\})$. The set $\{j : |i-j| < k\}$ has cardinality at most $2k$. Hence, with (5),

$$\#\left(\mathcal{J}_{w,S} \cap \{j : |i-j| \geq k\}\right) = \#\mathcal{J}_{w,S} + O(k) = \frac{|S|}{\mu_k(w)} + O(k).$$

We move on to the next sum. Using the sum of the geometric series with $\sigma < 1$, we get the bound

$$\sum_{i\in\mathcal{J}_{w,s}} \sum_{\substack{j\in\mathcal{J}_{w,S}\\|i-j|\geq k}} \sigma^{|i-j|-k} \leq \#\mathcal{J}_{w,S} \sum_{n=-\infty}^{\infty} \sigma^{|n|} = O\left(1/\mu_k(w)\right).$$

The contraction ratio property yields

$$\begin{aligned}
E_3 &= \mu_k^2(w)\left(\left(\frac{|S|}{\mu_k(w)} + O(1)\right)\left(\frac{|S|}{\mu_k(w)} + O(k)\right) + O\left(1/\mu_k(w)\right)\right)\\
&= |S|^2 + O\left(k\mu_k(w)\right)\\
&= |S|^2 + O\left(k\rho^k\right).
\end{aligned}$$

Since $\mathbb{E}_\mu[(M_k^w(S))^2] = E_1 + E_2 + E_3$, the above estimates complete the proof of this point.

Point 3. By definition,

$$\begin{aligned}
\mathbb{V}_\mu(M_k^w) &= \mathbb{E}_\mu\left[(M_k^w)^2\right] - \mathbb{E}_\mu\left[M_k^w\right]^2\\
&= |S| + |S|^2 + O(k\rho^k) + O\left(\sum_{\ell\in\pi_w} \rho^\ell\right) - \left(|S| + O(\rho^k)\right)^2\\
&= |S| + O(k\rho^k) + O\left(\sum_{\ell\in\pi_w} \rho^\ell\right).
\end{aligned}$$

□

Lemma 2. *Let μ be an invariant exponentially ψ-mixing measure on $\Omega^{\mathbb{N}}$ with contraction ratio μ. For each $k \in \mathbb{N}$, consider the projection μ_k of μ over Ω^k. Let $S \subset \mathbb{R}^+$ be a finite union of bounded intervals. Then, the following holds*

$$\mathbb{E}_{\mu\times\mu_k}[M_k(S)] = |S| + O(\rho^k).$$

Proof. By definition of $\mathbb{E}_{\mu\times\mu_k}[M_k(S)]$ and then Point 1 of Lemma 1 we obtain,

$$\mathbb{E}_{\mu\times\mu_k}[M_k(S)] = \sum_{w\in\Omega^k} \mu_k(w)\mathbb{E}_\mu[M_k^w(S)] = \sum_{w\in\Omega^k} \mu_k(w)\left(|S| + O(\rho^k)\right) = |S| + O(\rho^k).$$

□

The total variation distance d_{TV} between two probability measures Q and R on a sigma-algebra $\mathcal{F}$ is defined via

$$d_{TV}(Q,R) = \sup_{A\in\mathcal{F}} |Q(A) - R(A)|.$$

The total variation distance between two random variables X and Y taking values in $\mathbb{N}$ is simply

$$d_{TV}(X,Y) = \sup_{h:\mathbb{N}\to\mathbb{R},\ |h|\le 1} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]|.$$

For each $w \in \Omega^*$ and each $S \subseteq R^+$ we bound the total variation distance between the distribution of the random variable $M_k^w(S)$ and the Poisson distribution using Chen's result [7, Theorem 4.4], stated below as Proposition 2. It considers a sequence (finite or infinite) of random variables $X_1, X_2, X_3 \ldots$ with the following mixing condition:

Exponentially ϕ-mixing condition: Let i,j be natural numbers with $j > i$ and let $\mathcal{B}_{i,j}$ be the sigma-algebra generated by the random variables $X_i, ..., X_j$. With $\phi : \mathbb{N} \to \mathbb{R}$, $\phi(m) = e^{-\alpha m}$ for some $\alpha > 0$, for every for every $A \in \mathcal{B}_{1,i}$ and for every $B \in \mathcal{B}_{j,\infty}$,

$$\left|\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)} - \mathbb{P}(B)\right| \le \phi(j-i) \quad \text{for any } j > i. \tag{6}$$

Notice that our exponentially ψ-mixing condition (2) implies exponentially ϕ-mixing (6).

Proposition 2 ([7, Chen's Theorem 4.4])**.** *Let $X_1, \ldots, X_n$ be a sequence of identically distributed random variables taking values in $\{0,1\}$ and satisfying the exponentially ϕ-mixing condition* (6) *for some $\alpha > 0$. Define*

$$W = \sum_{i=1}^n X_i \quad \textit{and} \quad \lambda = \mathbb{E}[W].$$

Then, for any $h : \mathbb{N} \to \mathbb{R}$ with $|h| \le 1$ and $n \ge 3$,

$$|\mathbb{E}(h(W)) - P(\lambda,h)| < C_1(\alpha)\min(\lambda^{-1/2},1)\Big(\mathbb{V}(W) - \lambda + (\lambda+1)^2\frac{\log n}{n}\Big)$$

where $\mathbb{E}[W]$ and $\mathbb{V}[W]$ denote the expectation and variation of W respectively, and $C_1(\alpha)$ depends only on α.

Proposition 3 ([7, Lemma 5.2])**.** *For any real positive λ and t and for any $h : \mathbb{N} \to \mathbb{R}$ with $|h| \le 1$,*

$$|P(\lambda,h) - P(t,h)| \le 2|\lambda - t|.$$

Lemma 3. *Fix $S \subseteq \mathbb{R}^+$ a finite union of bounded intervals. For each $k \in \mathbb{N}$ and $w \in \Omega^k$ the quantity $n(k,w) = \#\mathcal{J}_w(S)$ satisfies*

$$\frac{\log(n(k,w))}{n(k,w)} = O(k\rho^k).$$

Proof of Lemma 3. We use identity (5). By the existence of a contraction ratio, there is a constant $K > 0$, for which

$$n(k,w) = \#\mathcal{J}_w(S) \geq \frac{|S|}{K\rho^k} - m$$

where m is the number of intervals of S. We are assuming that m does not depend on k. Since $\rho^k \to 0$ as $k \to \infty$, for k large enough,

$$\frac{|S|}{K\rho^k} - m \geq \frac{|S|}{2K\rho^k},$$

and $\log n/n$ is decreasing as $n \to \infty$,

$$\frac{\log(n(k,w))}{n(k,w)} \leq \frac{\log\left(|S|/(2K\rho^k)\right)}{|S|/(2K\rho^k)} = O(k\rho^k).$$

□

The following result extends the one in [2] to an alphabet with infinitely many symbols.

Lemma 4 (Total variation distance between $\mathbb{E}_\mu[h(M_k^w(S))]$ and $P(|S|, h)$)**.** *Let μ be an invariant and exponentially ψ-mixing measure and let $S \subset \mathbb{R}$ be a finite union of bounded intervals. There exists $k_0 \in \mathbb{N}$ and $C > 0$ such that for every $k \geq k_0$ and for every $w \in \Omega^k$, for any $h : \mathbb{N} \to \mathbb{R}$ such that $|h| \leq 1$,*

$$|\mathbb{E}_\mu[h(M_k^w(S))] - P(|S|, h)| \leq C\left(k\rho^k + \sum_{\ell\in\pi_w} \rho^\ell\right).$$

The constant C depends on the measure $|S|$ and the number of intervals of the set S.

Proof of Lemma 4. Since (2) implies (6), we apply Proposition 2. Fix the set $S \subset \mathbb{R}$. For each $k \in \mathbb{N}$ and $w \in \Omega^k$, we consider the random variables $M_k^w(S)$, $k \geq 1$. Since we assumed (3) the measure μ is invariant, so the variables $I_i(x,w)$ are identically distributed. Consider

$$M_k^w(S) = \sum_{i\in\mathcal{J}_{w,S}} I_i(x,w), \quad \lambda = \lambda(k,w) = \mathbb{E}_\mu\left[M_k^w(S)\right] \text{ and } n = n(k,w) = \#\mathcal{J}_{w,S}.$$

By Lemma 1, $\mathbb{E}_\mu[M_k^w] = |S| + O(\rho^k)$ and

$$|\mathbb{V}_\mu[M_k^w] - \mathbb{E}_\mu[M_k^w]| = O\left(k\rho^k + \sum_{\ell\in\pi_w}\rho^\ell\right).$$

Since the variables $I_i(x,w)$ are exponentially ψ-mixing, for k large enough, $n(k,w) \geq 3$. Using the bound on $\log(n(k,w))/n(k,w)$ in Lemma 3,

$$\begin{aligned}|\mathbb{E}_\mu\left[h(M_k^w)\right] - P(\lambda,h))| &\leq C_1(\alpha)\left(\mathbb{V}[M_k^w] - \lambda + (\lambda+1)^2\frac{\log n}{n}\right)\\ &= O\left(k\rho^k + \sum_{\ell\in\pi_w}\rho^\ell\right).\end{aligned}$$

And by Proposition 3,

$$\begin{aligned}|\mathbb{E}_\mu[h(M_k^w)] - P(|S|, h)| &\leq |\mathbb{E}_\mu\left[h(M_k^w)\right] - P(\lambda,h))| + |P(\lambda,h) - P(|S|, h)|\\ &= O\left(k\rho^k + \sum_{\ell\in\pi_w}\rho^\ell\right).\end{aligned}$$

□

Lemma 5. *Let ℓ and k be two natural numbers so that $1 \leq \ell < k$. Let W_k^ℓ be the set of words $w \in \Omega^k$ with period ℓ. Then,*

$$\mu_k(W_k^\ell) = O(\rho^{k-\ell}).$$

Proof of Lemma 5. For a word $v \in \Omega^\ell$ and an integer $k \geq 0$, define $w = \mathrm{ext}(v,k)$ as the unique word w in Ω^k such that

$$w_i = v_j \quad \text{if} \quad i \equiv j \mod \ell$$

for any $i,j \in \mathbb{N}$, $1 \leq i \leq k$ and $1 \leq j \leq \ell$.

With the bounded distortion property and using the contraction ratio, we have, for any $w \in W_k^\ell$ there exists $v \in \Omega^\ell$ such that $w = v\,\mathrm{ext}(v, k-\ell)$ and therefore

$$\mu_k(w) = \mu_k(v\ \mathrm{ext}(v,k-\ell)) \leq R\mu_\ell(v)\mu_{k-\ell}(\mathrm{ext}(v,k-\ell)) \leq RK\mu_\ell(v)\rho^{k-\ell}.$$

In order to bound $\mu_k(W_k^\ell)$, observe that there is a bijective function $f : W_k^\ell \to \Omega^\ell$ which is simply defined as $f(w) = w[1 \dots \ell]$. Then,

$$\mu_k(W_k^\ell) = \sum_{w \in W_k^\ell} \mu_k(w) = \sum_{v \in \Omega^\ell} \mu_k\left(v\ \mathrm{ext}(v,k-\ell)\right) \leq RK\rho^{k-\ell} \sum_{v \in \Omega^\ell} \mu_\ell(v) = RK\rho^{k-\ell}.$$

$\square$

Lemma 6 ($\mu \times \mu_k$-expectation on $M_k(S)$ in $\Omega^{\mathbb{N}} \times \Omega^k$)**.** *Let μ be an invariant and exponentially ψ-mixing measure on $\Omega^{\mathbb{N}}$ with a contraction ratio $\rho \in (0,1)$. For any set $S \subset \mathbb{R}$ which is a finite union of bounded intervals and for any $h : \mathbb{N} \to \mathbb{R}$ such that $|h| \leq 1$, the sequence of random variables $(M_k(S))_{k \geq 1}$ satisfies that*

$$|\mathbb{E}_{\mu\times\mu_k}[h(M_k(S))] - P(|S|,h)| = O(k\rho^k).$$

The constant hidden in the O-term depends on the measure $|S|$ and the number of intervals of the set S.

Proof of Lemma 6. We deal with the product measure $\mu \times \mu_k$ on $\Omega \times \Omega_k$. By definition,

$$\mathbb{E}_{\mu\times\mu_k}[h(M_k(S))](h) = \sum_{w \in \Omega^k} \mu_k(w)\mathbb{E}_\mu[h(M_k^w(S))].$$

So, writing $P(|S|,h) = \sum_{w \in \Omega^k} \mu(w)P(|S|,h)$ and using the triangular inequality, we have

$$|\mathbb{E}_{\mu\times\mu_k}[h(M_k(S))] - P(|S|,h)| \leq \sum_{w \in \Omega^k} \mu_k(w)\,|\mathbb{E}_\mu[h(M_k^w(S))] - P(|S|,h)|.$$

The bounds on $|\mathbb{E}[h(M^w(S))] - P(|S|,h)|$ given in Lemma 4 yield

$$|\mathbb{E}_{\mu\times\mu_k}[h(M_k(S))] - P(|S|,h)| = O(k\rho^k) + O\left(\sum_{w \in \Omega^k} \mu_k(w) \sum_{\ell \in \pi_w} \rho^\ell\right).$$

Now, the following holds, with the help of Lemma 5,

$$\sum_{w \in \Omega^k} \mu_k(w) \sum_{\ell \in \pi_w} \rho^\ell = \sum_{\ell=1}^{k-1} \sum_{w \in W_k^\ell} \mu_k(w)\rho^\ell = \sum_{\ell=1}^{k-1} \rho^\ell \mu_k\left(W_k^\ell\right) \leq O\left(\sum_{\ell=1}^{k-1} \rho^\ell \rho^{k-\ell}\right) = O(k\rho^k).$$

Finally,

$$|\mathbb{E}_{\mu\times\mu_k}[h(M_k(S))] - P(|S|,h)| = O(k\rho^k).$$

$\square$

Lemma 7 (The annealed result)**.** *Let μ an invariant and exponentially ψ-mixing measure on $\Omega^{\mathbb{N}}$. The sequence of random measures $(M_k(.))_{k\geq 1}$ on $\mathbb{R}^+$ converges in distribution to a Poisson point process on $\mathbb{R}^+$ as k goes to infinity. More precisely, for any Borel set $S \subseteq \mathbb{R}^+$,*

$$\mu \times \mu_k \left((x, w) \in \Omega^{\mathbb{N}} \times \Omega^k : M_k(x, w)(S) = j\right) \to p(|S|, j) \text{ as } k \to \infty.$$

Proof of Lemma 7. To prove it we apply Kallenberg's result stated in Proposition 1 for the product measure $\mu \times \mu_k$ on $\Omega \times \Omega_k$. In order to prove the first condition in Kallenberg's Theorem, we remark that the expectation of the Poisson distribution of parameter $|S|$ is $|S|$. On the other hand, the expectation of $M_k(S)$ is displayed in Lemma 2: $\mathbb{E}_{\mu\times\mu_k}[M_k(S)] = |S| + O(\rho^k)$, which implies that

$$\limsup_{k\to\infty} \mathbb{E}_{\mu\times\mu_k}[M_k(S)] \leq |S|.$$

Now, we prove the second condition of Proposition 1. Lemma 6, specified on $h = \mathbb{1}_{\{0\}}$, implies that

$$\lim_{k\to\infty} \mathbb{E}_{\mu\times\mu_k}\left[\mathbb{1}_{\{0\}}(M_k(S))\right] = p(|S|, 0) = e^{-|S|}$$

Then, both conditions of Proposition 1 hold for every finite union of intervals S and we can conclude that $(M_k(.))_{k\geq 1}$ converges in distribution to a Poisson point process on $\mathbb{R}^+$. $\square$

2.2 The quenched result

Following a similar strategy as the one from [3, Proposition 3], we prove a concentration result to show that M_k^x converges to Poisson for μ-almost all $x \in \Omega^{\mathbb{N}}$. We need an inequality that holds for countable alphabets and functions depending on countably many variables with the bounded differences property known as the Lipschitz condition for a weighted Hamming distance, see Proposition 6. As mentioned in the Introduction, for this we adapt the martingale difference method given in [15, Theorem 3.3.1] and [16, Theorem 1.1]. The former deals with a finite alphabet and weighted Hamming distances and the latter deals with a countable alphabet but a constant Hamming distance.

We need to deal with infinitely many random variables $X_1, X_2, \ldots$ defined on a space $\mathbb{X}$ taking values on the countable alphabet Ω, and a function $\varphi : \Omega^{\mathbb{N}} \to \mathbb{R}$. We work with finitely many $X_1, \ldots, X_N$ and $\varphi_N : \Omega^N \to \mathbb{R}$ for N large and then take $N \to \infty$. For each N, consider the filtration

$$\mathcal{F}_N = \sigma(X_1, \ldots, X_N)$$

of sigma-algebras generated by the first N random variables and the Azuma–Hoeffding coefficients that are defined in terms of martingale differences.

Definition (Azuma–Hoeffding coefficients d_i)**.** *Let $X_1, X_2, X_3, \ldots$ be a sequence of random variables defined on $(\mathbb{X}, \mathcal{F}, \mathbb{P})$ taking values on some Ω. Consider the filtration of sub-sigma-algebras $\mathcal{F}_i$,*

$$\{\emptyset, \mathbb{X}\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}.$$

and $\mathcal{F}$ is the smallest sigma-algebra containing all the others. Consider a function $\varphi : \Omega^{\mathbb{N}} \to \mathbb{R}$. For $X = X_1 X_2 \cdots$, we define, for each $i \in \mathbb{N}$,

$$V_i(\varphi) = \mathbb{E}[\varphi(X) \mid \mathcal{F}_i] - \mathbb{E}[\varphi(X) \mid \mathcal{F}_{i-1}].$$

and we define the Azuma–Hoeffding coefficients

$$d_i = \sup_{\mathbb{X}} |V_i(\varphi)|. \tag{7}$$

We follow [15, page 23] and we the use η-mixing coefficients introduced there.

Definition (η-mixing coefficients and their associated matrix Δ)**.** *For a sequence $X_1, X_2, \ldots$ of random variables taking values a finite or countable alphabet Ω and for any $i, j \in \mathbb{N}$, we write $X^{\geq j}$ and $X^{\leq i}$ for $(X_j, X_{j+1}, \ldots)$ and $(X_1, \ldots, X_i)$ respectively. The mixing coefficients $\eta_{i,j}$ are real numbers such that, for $j \geq i$,*

$$\eta_{ij} = \sup_{\substack{x,x' \in \Omega^i, \\ x[i] \neq x'[i]}} \sup_{A \in \sigma(X^{\geq j})} \left| \mathbb{P}\left(X^{\geq j} \in A | X^{\leq i} = x\right) - \mathbb{P}\left(X^{\geq j} \in A | X^{\leq i} = x'\right) \right|$$

that is, the supremum is taken over x and x' which are elements of Ω^i which differ only in the i^{th} coordinate and $\sigma(X^{\geq j})$ is the sigma-algebra of Borel sets not depending the first $j-1$ coordinates. If $\mathbb{P}\left[X^{\leq i} = x\right] = 0$ or $\mathbb{P}\left[X^{\leq i} = x'\right] = 0$, we define $\eta_{ij} = 0$. The corresponding matrix Δ is defined as

$$\Delta_{ij} = \begin{cases} \eta_{ij}, & \text{if } j > i \\ 1, & \text{if } j = i \\ 0, & \text{if } j < i. \end{cases}$$

We consider the Banach space $(\ell^2, \| \; \|)$ which is formed by all the real sequences

$$v = (v_j)_{j \in \mathbb{N}} \quad \text{so that} \quad \|v\|^2 = \sum_{j \geq 1} v_j^2 < \infty.$$

The matrix Δ induces a bounded linear operator $\Delta : \ell^2 \to \ell^2$ whose norm is defined as

$$\|\Delta\| = \sup_{\|v\|=1} \|\Delta v\|.$$

The functions we are going to consider in the concentration inequalities below satisfy a bounded differences property known as Lipschitz condition for a weighted Hamming distance.

Definition (c-Lipschitz)**.** *Given $c = (c_i)_{i \geq 1}$, with $c_i \in \mathbb{R}$, we say that the function $\varphi : \Omega^{\mathbb{N}} \to \mathbb{R}$ is c-Lipschitz if for any $x, x' \in \Omega^{\mathbb{N}}$ which differ only in coordinate i, we have,*

$$|\varphi(x) - \varphi(x')| \leq c_i.$$

Proposition 4 (Upper bound for the Azuma–Hoeffding coefficients)**.** *Let $X_1, X_2, \ldots$ a sequence of random variables defined on $(\mathbb{X}, \mathcal{F}, \mathbb{P})$ taking values on alphabet Ω. Let Δ be the matrix of η-mixing coefficients with $\|\Delta\| < \infty$. Let $c = (c_i)_{i \geq 1} \in \ell^2$ and $\varphi : \Omega^{\mathbb{N}} \to \mathbb{R}$ a bounded function that is also c-Lipschitz. Then, the Azuma–Hoeffding coefficients $d = (d_i)_{i \geq 1}$ satisfy, for each $i \geq 1$,*

$$0 \leq d_i \leq \sum_{j \geq i} c_j \eta_{ij}. \tag{8}$$

In particular, $\|d\| \leq \|\Delta\| \|c\|$.

Proof. For each i, consider the filtration $\mathcal{F}_i = \sigma(X_1, \ldots, X_i)$ of sigma-algebras generated by the first i random variables. We want to bound the Azuma–Hoeffding coefficients as in (7) We prove (8) for $i = 1$, the others are similar. For $x \in \Omega^N$ we denote $\mathbb{E}[\varphi | X = x]$ by $\mathbb{E}[\varphi | x]$ and similarly for probabilities. Let us fix $x \in \Omega^N$ so that $\mathbb{P}[x^{\leq 1}] > 0$. Then,

$$\begin{aligned} V_1(\varphi) =& \mathbb{E}\left[\varphi | x^{\leq 1}\right] - \mathbb{E}[\varphi] \\ &= \int_{y \in \Omega^{\mathbb{N}}} \mathbb{E}\left[\varphi | x^{\leq 1} y^{>1}\right] \mathrm{d}\mathbb{P}\left(x^{\leq 1} y^{>1} | x^{\leq 1}\right) - \mathbb{E}[\varphi | y] \, \mathrm{d}\mathbb{P}(y) \\ &= \int_{y \in \Omega^{\mathbb{N}}} \left(\mathbb{E}\left[\varphi | x^{\leq 1} y^{>1}\right] - \mathbb{E}[\varphi | y]\right) \mathrm{d}\mathbb{P}(y) + \int_{y \in \Omega^{\mathbb{N}}} \mathbb{E}\left[\varphi | x^{\leq 1} y^{>1}\right] \left(\mathrm{d}\mathbb{P}\left(x^{\leq 1} y^{>1} | x^{\leq 1}\right) - \mathrm{d}\mathbb{P}(y)\right). \end{aligned}$$

Since $1 = \int_{y\in\Omega^{\mathbb{N}}} \mathrm{d}\mathbb{P}\left(x^{\leq 1}y^{>1}|x^{\leq 1}\right) = \int_{y\in\Omega^{\mathbb{N}}} \mathrm{d}\mathbb{P}\,(y)$,

$$V_1(\varphi) = \int_{y\in\Omega^{\mathbb{N}}} \left(\mathbb{E}\left[\varphi|x^{\leq 1}y^{>1}\right] - \mathbb{E}\left[\varphi|y^{\leq 1}y^{>1}\right]\right)\mathrm{d}\mathbb{P}\,(y)$$
$$+ \int_{y\in\Omega^{\mathbb{N}}} \left(\mathbb{E}\left[\varphi|x^{\leq 1}y^{>1}\right] - \mathbb{E}\left[\varphi|x\right]\right)\left(\mathrm{d}\mathbb{P}\left(x^{\leq 1}y^{>1}|x^{\leq 1}\right) - \mathrm{d}\mathbb{P}\,(y)\right),$$

telescoping, for each $N \in \mathbb{N}$ and $N \geq 2$,

$$= \int_{y\in\Omega^{\mathbb{N}}} \left(\mathbb{E}\left[\varphi|x^{\leq 1}y^{>1}\right] - \mathbb{E}\left[\varphi|y^{\leq 1}y^{>1}\right]\right)\mathrm{d}\mathbb{P}\,(y)$$
$$+ \int_{y\in\Omega^{\mathbb{N}}} \left(\sum_{k=2}^{N} \mathbb{E}\left[\varphi|x^{<k}y^{\geq k}\right] - \mathbb{E}\left[\varphi|x^{\leq k}y^{>k}\right]\right)\left(\mathrm{d}\mathbb{P}\left(x^{\leq 1}y^{>1}|x^{\leq 1}\right) - \mathrm{d}\mathbb{P}\,(y)\right)$$
$$+ \int_{y\in\Omega^{\mathbb{N}}} \mathbb{E}\left[\varphi|x^{\leq N}y^{>N}\right]\left(\mathrm{d}\mathbb{P}\left(x^{\leq 1}y^{>1}|x^{\leq 1}\right) - \mathrm{d}\mathbb{P}\,(y)\right)$$

interchanging sum with integral

$$= \int_{y\in\Omega^{\mathbb{N}}} \left(\mathbb{E}\left[\varphi|x^{\leq 1}y^{>1}\right] - \mathbb{E}\left[\varphi|y^{\leq 1}y^{>1}\right]\right)\mathrm{d}\mathbb{P}\,(y)$$
$$+ \sum_{k=2}^{N} \int_{y\in\Omega^{\mathbb{N}}} \left(\mathbb{E}\left[\varphi|x^{<k}y^{\geq k}\right] - \mathbb{E}\left[\varphi|x^{\leq k}y^{>k}\right]\right)\left(\mathrm{d}\mathbb{P}\left(x^{\leq 1}y^{>1}|x^{\leq 1}\right) - \mathrm{d}\mathbb{P}\,(y)\right)$$
$$+ \int_{y\in\Omega^{\mathbb{N}}} \mathbb{E}\left[\varphi|x^{\leq N}y^{>N}\right]\left(\mathrm{d}\mathbb{P}\left(x^{\leq 1}y^{>1}|x^{\leq 1}\right) - \mathrm{d}\mathbb{P}\,(y)\right)$$

$$= \int_{y\in\Omega^{\mathbb{N}}} \left(\mathbb{E}\left[\varphi|x^{\leq 1}y^{>1}\right] - \mathbb{E}\left[\varphi|y^{\leq 1}y^{>1}\right]\right)\mathrm{d}\mathbb{P}\,(y)$$
$$+ \sum_{k=2}^{N} \int_{\{y^{\geq k}\in\Omega^{\mathbb{N}}\}\cap\{y^{<N}\in\Omega^{\mathbb{N}}\}} \left(\mathbb{E}\left[\varphi|x^{<k}y^{\geq k}\right] - \mathbb{E}\left[\varphi|x^{\leq k}y^{>k}\right]\right) \int_{y^{<k}\in\Omega^{k-1}} \left(\mathrm{d}\mathbb{P}\left(x^{\leq 1}y^{\geq 1}|x^{\leq 1}\right) - \mathrm{d}\mathbb{P}\,(y)\right)$$
$$+ \int_{y\in\Omega^{\mathbb{N}}} \mathbb{E}\left[\varphi|x^{\leq N}y^{>N}\right]\left(\mathrm{d}\mathbb{P}\left(x^{\leq 1}y^{>N}|x^{\leq 1}\right) - \mathrm{d}\mathbb{P}\,(y)\,.\right)$$

For fixed $k = 2, \ldots, N$ we have

$$\int_{y^{<k}\in\Omega^{k-1}} \left(\mathrm{d}\mathbb{P}\left(x^{\leq 1}y^{\geq 1}|x^{\leq 1}\right) - \mathrm{d}\mathbb{P}\left(y^{\geq 1}\right)\right) = \mathrm{d}(\mathbb{P}(x^{\leq 1}y^{\geq k}|x^{\leq 1}) - \mathrm{d}(\mathbb{P}(y^{\geq k}).$$

Then,

$$V_1(\varphi) = \int_{y\in\Omega^{\mathbb{N}}} \left(\mathbb{E}\left[\varphi|x^{\leq 1}y^{>1}\right] - \mathbb{E}\left[\varphi|y^{\leq 1}y^{>1}\right]\right)\mathrm{d}\mathbb{P}\,(y)$$
$$+ \sum_{k=2}^{N} \int_{\{y^{\geq k}\in\Omega^{\mathbb{N}}\}\cap\{y^{<N}\in\Omega^{\mathbb{N}}\}} \left(\mathbb{E}\left[\varphi|x^{<k}y^{\geq k}\right] - \mathbb{E}\left[\varphi|x^{\leq k}y^{>k}\right]\right)\left(\mathrm{d}\mathbb{P}\left(x^{\leq 1}y^{\geq k}|x^{\leq 1}\right) - \mathrm{d}\mathbb{P}\left(y^{\geq k}\right)\right)$$
$$+ \int_{y\in\Omega^{\mathbb{N}}} \mathbb{E}\left[\varphi|x^{\leq N}y^{>N}\right]\left(\mathrm{d}\mathbb{P}\left(x^{\leq 1}y^{>N}|x^{\leq 1}\right) - \mathrm{d}\mathbb{P}\,(y)\,,\right)$$
$$\leq \sum_{j=1}^{N} c_j\eta_{1j} + \sup_{\Omega^{\mathbb{N}}}|\varphi|\cdot\eta_{1,N+1}.$$

Since $||\Delta|| < \infty$, $\eta_{1,N+1}$ goes to zero when $N \to \infty$, and $d_1 \leq \sum_{j=1} c_j\eta_{1j}$. The proof is then complete. $\square$

Proposition 5 (Instantiation of [18, Lemma 4.1])**.** *Let $X = X_1, X_2, \ldots, X_N$ be a sequence of random variables defined on $(\mathbb{X}, \mathcal{F}, \mathbb{P})$ taking values on some Ω. For $1 \le i \le N$, consider the Azuma–Hoeffding coefficients d_i as in* (7)*. Then,*

$$\mathbb{P}\left(|\varphi(X) - \mathbb{E}[\varphi(X)]| \ge t\right) \le 2\exp\left(\frac{-t^2}{2\sum_{i=1}^N d_i^2}\right).$$

Proposition 6 (Concentration inequality with infinitely many variables)**.** *Let $X_1, X_2, \ldots$ be random variables taking values in some countable set Ω and let Δ be the matrix of mixing coefficients with $\|\Delta\| < \infty$. Let $\varphi : \Omega^{\mathbb{N}} \to \mathbb{R}$ be a c-Lipschitz function for some $c = (c_i)_{i\ge1} \in \ell^2$ such that $\varphi(X) \in L^1$. Then, for any $t > 0$,*

$$\mathbb{P}\left(|\varphi(X) - \mathbb{E}[\varphi(X)]| \ge t\right) \le 2\exp\left(\frac{-t^2}{2\|\Delta\|^2\|c\|^2}\right).$$

Proof. We assume first that φ is bounded. For each $N \in \mathbb{N}$ let $\varphi : \Omega^{\mathbb{N}} \to \mathbb{R}$,

$$\varphi_N(x) = \mathbb{E}\left[\varphi(X) \,|\, X^{\le N} = x^{\le N}\right].$$

In other words, φ_N is the martingale $\mathbb{E}[\varphi|\mathcal{F}_N]$. Since φ is bounded, by Lévy's zero-one law the martingale $\mathbb{E}[\varphi|\mathcal{F}_N]$ converges to $\varphi(X)$ both a.e. and in L^1. By standard properties of conditional expectations, it is easy to see that

$$V_i(\varphi_N) = \begin{cases} V_i(\varphi) & \text{if } i \le N \\ 0 & \text{if } i > N. \end{cases}$$

Hence, the Azuma–Hoeffding coefficients d_i for φ agree with those for φ_N as long as $i \le N$ and vanish afterwards.

We invoke Proposition 5 with φ_N and get

$$\mathbb{P}\left(|\varphi_N(X) - \mathbb{E}[\varphi_N(X)]| \ge t\right) \le 2\exp\left(\frac{-t^2}{2\sum_{i=1}^N d_i^2}\right).$$

By Proposition 4 (keep in mind we are assuming φ bounded now) this becomes

$$\mathbb{P}\left(|\varphi_N(X) - \mathbb{E}[\varphi_N(X)]| \ge t\right) \le 2\exp\left(\frac{-t^2}{2\|\Delta\|^2\|c\|^2}\right).$$

The proof for the case φ being bounded finishes by taking the limit as $N \to \infty$.

Now remove the boundedness hypothesis on φ. By dominated convergence, the sequence of bounded functions

$$\varphi_N = \begin{cases} N, & \text{if } \varphi(x) > N \\ \varphi(x), & \text{if } |\varphi(x)| \le N \\ -N, & \text{if } \varphi(x) < -N \end{cases}$$

converges to φ both a.e. and in L^1. It remains to check that these φ_N's are still $c-$Lipschitz. We can think of the truncations as $\varphi_N = \Psi_N \circ \varphi$, with

$$\Psi_N(x) = \frac{|x+N| - |x-N|}{2} = \begin{cases} N, & \text{if } x > N \\ x, & \text{if } |x| \le N \\ -N, & \text{if } x < -N \end{cases}$$

having Lipschitz constant 1. Therefore, for every pair $x, x' \in \Omega^{\mathbb{N}}$ differing only in the ith coordinate we have

$$|\varphi_N(x) - \varphi_N(x')| = |\Psi_N(\varphi(x)) - \Psi_N(\varphi(x'))| \le |\varphi(x) - \varphi(x')| \le c_i.$$

This completes the proof. $\square$

The next results consider our particular case of $\mathbb{X} = \Omega^{\mathbb{N}}$, the invariant mixing measure μ, the usual projections $X_i : \Omega^{\mathbb{N}} \to \Omega$ onto the i-th coordinate of $x \in \Omega^{\mathbb{N}}$ and $\mathcal{F}_i = \mathcal{B}_{1,i}$ for all $i \in \mathbb{N}$.

Lemma 8 (Upper bound for $\|\Delta\|$)**.** *Let μ an invariant exponentially ψ-mixing measure on $\Omega^{\mathbb{N}}$ with mixing constants $T > 0$ and $\sigma \in (0,1)$. Let Δ be the matrix of η-mixing coefficients. Then,*

$$\|\Delta\| \leq 1 + 2T\sigma/(1-\sigma).$$

Proof. Recall that we write $X^{\geq j}$ and $X^{\leq i}$ for $(X_j, X_{j+1}, \ldots)$ and $(X_1, \ldots, X_i)$ respectively. Thus, for $w \in \Omega^i$, the probability that $X^{\leq i} = w$ is $\mu_i(w)$. For $i, j \in \mathbb{N}$ such that $j \geq i$, let $x, x' \in \Omega^{\mathbb{N}}$ which differ only in coordinate i, and let $A \in \mathcal{B}_{j,\infty}$ where $j > i$. Let us recall that

$$\eta_{i,j} = \sup_{\substack{x,x' \in \Omega^i,\\ x[i] \neq x'[i]}} \sup_{A \in \mathcal{B}_{j,\infty}} \left| \frac{\mu(C_i(x) \cap A)}{\mu_i(x)} - \frac{\mu(C_i(x') \cap A)}{\mu_i(x')} \right|.$$

By our mixing property (2),

$$\left| \frac{\mu(C_i(x) \cap A)}{\mu_i(x)} - \mu(A) \right| + \left| \mu(A) - \frac{\mu(C_i(x') \cap A)}{\mu_i(x')} \right| \leq 2\mu(A) T \sigma^{j-i}.$$

Then, each η-mixing coefficient $\eta_{ij} \leq 2T\sigma^{j-i}$.

Let J be the matrix

$$J_{ij} = \begin{cases} 1 & \textit{if } j = i+1 \\ 0 & \textit{otherwise}, \end{cases}$$

and observe that $\|J\| \leq 1$. By the triangular inequality and geometric series we get

$$\|\Delta\| \leq 1 + T \sum_{i \geq 1} \|(\sigma J)^i\| \leq 1 + 2\frac{T\sigma}{1-\sigma}.$$

□

Lemma 9 (Two concentrations)**.** *Let μ an invariant and exponentially ψ-mixing measure on $\Omega^{\mathbb{N}}$. Let $k \in \mathbb{N}$ be large enough, let $j \geq 0$, let $S \subset \mathbb{R}$ be a finite union of bounded intervals and denote by $\sup S$ the supremum of S.*

1. *For $\varphi_{k,S} : \Omega^{\mathbb{N}} \to \mathbb{R}$, $\varphi_{k,S}(x) = \mathbb{E}_{\mu_k}[M_k^x(S)]$ we have, for any $t > 0$,*

$$\mu\left(\{x \in \Omega^{\mathbb{N}} : |\varphi_{k,S}(x) - \mathbb{E}_\mu[\varphi_{k,S}]| \geq t\}\right) \leq 2\exp\left(\frac{-t^2}{\|\Delta\|^2 8k^4 (\sup S) K^2 \rho^k}\right).$$

2. *For $\varphi_{k,j,S} : \Omega^{\mathbb{N}} \to [0,1]$, $\varphi_{k,j,S}(x) = \mu_k(\{w \in \Omega^k : M_k^x(w)(S) = j\})$ we have, for any $t > 0$,*

$$\mu\left(\{x \in \Omega^{\mathbb{N}} : |\varphi_{k,j,S}(x) - \mathbb{E}_\mu[\varphi_{k,j,S}]| \geq t\}\right) \leq 2\exp\left(\frac{-t^2}{\|\Delta\|^2 8k^2 (\sup S) K^2 \rho^k}\right).$$

The numbers $\rho \in (0,1)$, and $K > 0$ are the contraction ratio and the constant given in (4); and Δ is the matrix of η-mixing coefficients associated with the mixing measure μ.

Proof. Define the sequence of random variables $X_i : \Omega^{\mathbb{N}} \to \Omega$ as the usual projection onto the i-th coordinate of $x \in \Omega^{\mathbb{N}}$. Fix $k \in \mathbb{N}$ and fix a set S which is a finite union of bounded intervals. For $i_0 \in \mathbb{N}$ and $x, x' \in \Omega^{\mathbb{N}}$ such that x and x' differ only in the i_0-th coordinate, we define the set

$$D_{k,i_0,S}(x,x') = \{w \in \Omega^k : M_k^x(w)(S) \neq M_k^{x'}(w)(S)\}.$$

Since x and x' only differ in the coordinate i_0, a word $w \in D_{k,i_0,S}(x,x')$ is one of the following: $x[i_0-k+i, i_0+i)$ or $x'[i_0-k+i, i_0+i)$ for $i \in \mathbb{N}$ and $1 \le i \le k$ (with the convention that if $i < 0$, the word $x[i, i+k)$ is the empty word). Thus, there are, at most, $2k$ words in $D_{k,i_0,S}(x,x')$. For each $w \in D_{k,i_0,S}(x,x')$,

$$\mu_k(w) \le \sup S / i_0$$

because the sum in the definition of $M_k^x(w)(S)$ runs over the indexes i in

$$\mathcal{J}_{w,S} = \{i \in \mathbb{N} : i\mu_k(w) \in S\}.$$

On the other hand, the measure of every word $w \in \Omega^k$ is upper bounded by the contraction ratio ρ, $0 < \rho < 1$,

$$\mu_k(w) \le K\rho^k.$$

Point 1. Let $\varphi_{k,S} : \Omega^{\mathbb{N}} \to \mathbb{R}$, $\varphi_{k,S}(x) = \mathbb{E}_{\mu_k}[M_k^x(S)]$. We prove that there exists $(c_i)_{i\ge1}$ in ℓ^2 such that, for every N, $\varphi_{k,S}(x)$ is c-Lipchitz. In fact,

$$|\varphi_{k,S}(x) - \varphi_{k,S}(x')| \le \mu_k(D_{k,i_0,S}(x,x')) \sup_{w \in D_{k,i_0,S}(x,x')} \left| M_k^x(w)(S) - M_k^{x'}(w)(S) \right|.$$

If x and x' differ only in coordinate i_0, for each $w \in \Omega^k$, $M_k^x(w)(S)$ and $M_k^{x'}(w)(S)$ differ, at most, in the values of the k indicators $I_{i+i_0-k}(\cdot, w)$ for $1 \le i \le k$. This implies that $|M_k^x(w)(S) - M_k^{x'}(w)(S)| \le k$. Define $c = (c_i)_{i\in\mathbb{N}}$,

$$c_i = 2k^2 \min(K\rho^k, \sup S/i).$$

Then,

$$|\varphi_{k,S}(x) - \varphi_{k,S}(x')| \le c_i,$$

which means that $\varphi_{k,S}$ is c-Lipchitz. Let's see that $(c_i)_{i\ge1}$ is in ℓ^2:

$$\begin{aligned}
\|c\|^2 = \sum_{i\ge1} c_i^2 &= \sum_{i \le \sup S/(K\rho^k)} c_i^2 + \sum_{i > \sup S/(K\rho^k)} c_i^2 \\
&\le 4k^2 \left(K^2 \sum_{i \le \sup S/(K\rho^k)} \rho^{2k} + \sup S^2 \sum_{i > \sup S/(K\rho^k)} 1/i^2 \right) \\
&\le 4k^2 (\sup S) K^2 \rho^k.
\end{aligned}$$

The function $\varphi_{k,S}$ is in $L^1(\mu)$ because of Lemma 2 and thus it satisfies the assumptions of Proposition 6. Then,

$$\mu(\{x \in \Omega^{\mathbb{N}} : |\varphi_{k,S}(x) - \mathbb{E}_\mu[\varphi_{k,S}]| \ge t\} \le 2\exp\left(\frac{-t^2}{2\|\Delta\|^2\|c\|^2}\right).$$

Point 2. Fix $j \ge 0$. Let $\varphi_{k,j,S} : \Omega^{\mathbb{N}} \to [0,1]$, $\varphi_{k,j,S}(x) = \mu_k(\{w \in \Omega^k : M_k^x(w)(S) = j\})$. Define $c = (c_i)_{i\in\mathbb{N}}$ as

$$c_i = 2k \min(K\rho^k, \sup S/i).$$

With the same argument as in Point 1, $c = (c_i)_{i\in\mathbb{N}}$ is in ℓ^2 and for any pair $x, x' \in \Omega^{\mathbb{N}}$ which only differ in the ith-coordinate we have, for each $i \ge 1$,

$$|\varphi_{k,j,S}(x) - \varphi_{k,j,S}(x')| \le \mu_k(D_{k,i,S}(x,x')) \le c_i.$$

Hence, $\varphi_{k,j,S}$ is c-Lipchitz. It is also bounded and hence it is in $L^1(\mu)$. The assumptions of Proposition 6 are fulfilled. Then,

$$\mu(\{x \in \Omega^{\mathbb{N}} : |\varphi_{k,j,S}(x) - \mathbb{E}_\mu[\varphi_{k,j,S}]| \geq t\} \leq 2 \exp\left(\frac{-t^2}{2\|\Delta\|^2\|c\|^2}\right).$$

□

The next lemma gives the wanted quenched result: it proves that for μ-almost all $x \in \Omega^{\mathbb{N}}$, the sequence of random measures $(M_k^x(.))_{k\geq 1}$ on $\mathbb{R}^+$ converges in distribution to a Poisson point process on $\mathbb{R}^+$ as k goes to infinity.

Lemma 10 (The quenched result)**.** *Let μ an invariant and exponentially ψ-mixing measure on $\Omega^{\mathbb{N}}$. Then for μ-almost all $x \in \Omega^{\mathbb{N}}$, the sequence of random measures $(M_k^x(.))_{k\geq 1}$ on $\mathbb{R}^+$ converges in distribution to a Poisson point process on $\mathbb{R}^+$ as k goes to infinity.*

Proof. To prove it we apply Kallenberg's result stated in Proposition 1 for the measure μ. Fix $S \subseteq \mathbb{R}^+$ a finite union of intervals with rational endpoints. Let $j \geq 0$ be an integer. We need to show that for μ-almost all $x \in \Omega^{\mathbb{N}}$,

1. $\limsup_{k\to\infty} \mathbb{E}_{\mu_k}[M_k^x(S)] \leq |S|$.
2. $\lim_{k\to\infty} \mu_k(\{w \in \Omega^k : M_k^x(w)(S) = j\}) = p(|S|, j)$.

We start with Point 1. Let $\varphi_{k,S} : \Omega^{\mathbb{N}} \to \mathbb{R}$, $\varphi_{k,S}(x) = \mathbb{E}_{\mu_k}[M_k^x(S)]$ By Point 1 of Lemma 9 we have, for any $t > 0$,

$$\mu\left(\{x \in \Omega^{\mathbb{N}} : |\varphi_{k,S}(x) - \mathbb{E}_\mu[\varphi_{k,S}]| \geq t\}\right) \leq 2 \exp\left(\frac{-t^2}{\|\Delta\|^2 8k^4(\sup S)K^2\rho^k}\right).$$

Taking $t_k = 1/k$, using the bound for Δ given in Lemma 8, and using that $1/(k^6\rho^k)$ goes to infinity when k goes to infinity, this series converges

$$\sum_{k\geq 1} \mu(|\varphi_{k,S} - \mathbb{E}_\mu[\varphi_{k,S}]| > t_k) < \infty.$$

By the Borel–Cantelli lemma the limsup event

$$\left\{\, x \in \Omega^{\mathbb{N}} : \, |\varphi_{k,S}(x) - \mathbb{E}_\mu[\varphi_{k,S}(x)]| > t_k \text{ for infinitely many } k\right\}$$

has μ-probability zero. Then, there exists a set of μ-measure 1 (which may depend on S) where the difference

$$\varphi_{k,S}(x) - \mathbb{E}_\mu[\varphi_{k,S}(x)]$$

goes to zero as k goes to infinity . Since

$$\mathbb{E}_\mu[\varphi_{k,S}(x)] = \mathbb{E}_{\mu\times\mu_k}[M_k^x(S)]$$

and, by Lemma 2,

$$\mathbb{E}_{\mu\times\mu_k}[M_k^x(S)] = |S| + O(\rho^k),$$

it follows that $\varphi_{k,S}(x) = \mathbb{E}_{\mu_k}[M_k^x(S)]$ converges to $|S|$ as k goes to infinity in a set of μ-measure 1. Hence, $\limsup_{k\to\infty} \mathbb{E}_{\mu_k}[M_k^x(S)] \leq |S|$.

Point 2. By Lemma 7, for every S that is a finite union of intervals with rational endpoints and for every integer $j \geq 0$,

$$\mu \times \mu_k\left((x, w) \in \Omega^{\mathbb{N}} \times \Omega^k : M_k(x, w)(S) = j\right) \to p(|S|, j) \text{ as } k \to \infty.$$

Using that for any Borel set $A \subseteq \mathbb{R}^+$ the equality $\mathbb{E}_\mu[\mu_k(A)] = \mu \times \mu_k(A)$ holds, we have

$$\mathbb{E}_\mu \left[\mu^k \left(w \in \Omega^k : M_k^x(S)(w) = j \right) \right] \to p(|S|, j) \text{ as } k \to \infty.$$

Let $\varphi_{k,j,S} : \Omega^{\mathbb{N}} \to (0,1)$, $\varphi_{k,j,S}(x) = \mu_k(\{w \in \Omega^k : M_k^x(w)(S) = j\})$. By Point 2 of Lemma 9,

$$\mu \left(\{x \in \Omega^{\mathbb{N}} : |\varphi_{k,j,S}(x) - \mathbb{E}_\mu[\varphi_{k,j,S}]| \geq t\} \right) \leq 2 \exp \left(\frac{-t^2}{\|\Delta\|^2 8k^2 (\sup S) K^2 \rho^k} \right).$$

Taking $t_k = 1/k$ and using that $1/(k^4 \rho^k)$ goes to infinity when k goes to infinity, this series converges

$$\sum_{k \geq 1} \mu(|\varphi_{k,j,S} - \mathbb{E}_\mu[\varphi_{k,j,S}]| > t_k) < \infty.$$

By the Borel–Cantelli lemma the limsup event

$$\left\{ x \in \Omega^{\mathbb{N}} : |\varphi_{k,j,S}(x) - \mathbb{E}_\mu[\varphi_{k,j,S}(x)]| > t_k \text{ for infinitely many } k \right\}$$

has μ-probability zero. Then, there exists a set of μ-measure 1 (which may depend on S) where the difference

$$\varphi_{k,j,S}(x) - \mathbb{E}_\mu[\varphi_{k,j,S}(x)]$$

goes to zero as k goes to infinity . Since $\mathbb{E}_\mu \left[\mu^k \left(w \in \Omega^k : M_k^x(S)(w) = j \right) \right]$ converges to $p(|S|, j)$ as k goes to infinity in a set of μ-measure 1, for μ-almost all $x \in \Omega^{\mathbb{N}}$,

$$\left| \mu_k \left(\{w \in \Omega^k : M_k^x(S)(w) = j\} \right) - p(|S|, j) \right| \to 0 \quad \text{as } k \to \infty.$$

There are countably many sets S that are finite union of intervals with rational endpoints, and there are countably many integer values $j \geq 0$. Then there are just a countable number of each such exceptional sets (because their countable union has also μ-measure zero. Using Proposition 1 we conclude that for μ-almost every $x \in \Omega^{\mathbb{N}}$, $M_k^x(.)$ converges in distribution to a Poisson point process on $\mathbb{R}^+$. □

Lemma 10 shows that μ-almost all $x \in \Omega^{\mathbb{N}}$ are Poisson-generic. Theorem 1 is proved.

Acknowledgements

We are grateful to Inés Armendariz and Frédéric Paccaut. This work is supported by grants CONICET PIP 11220210100220CO, UBACyT 20020220100065BA, PICTO UNGS-00001 (2021), and UNGS-IDH 30/3373.

References

[1] Miguel Abadi. Exponential approximation for hitting times in mixing processes. *Mathematical Physics Electronic Journal*, 7:Paper No. 2, 19 p., 2001.

[2] Miguel Abadi and Nicolas Vergne. Sharp error terms for return time statistics under mixing conditions. *J. Theor. Probab.*, 22(1):18–37, 2009.

[3] Nicolás Alvarez, Verónica Becher, and Martín Mereb. Poisson generic sequences. *International Mathematics Research Notices*, rnac234, 2022.

[4] Guy Barat, Valérie Berthé, Pierre Liardet, and Jörg Thuswaldner. Dynamical directions in numeration. *Ann. Inst. Fourier (Grenoble)*, 56(7):1987–2092, 2006.

[5] Verónica Becher and Gabriel Sac Himelfarb. Construction of a λ-Poisson generic sequence. *Mathematics of Computation*, 92:1453–1466, 2023.

[6] Valérie Berthé and Jungwon Lee. *Trans. Amer. Math. Soc.*, 376:7947–7982, 2023.

[7] Louis H. Y. Chen. Poisson Approximation for Dependent Trials. *The Annals of Probability*, 3(3):534 – 545, 1975.

[8] W. Doeblin. Remarques sur la théorie métrique des fractions continues. *Compos. Math.*, 7:353–371, 1940.

[9] Anish Ghosh, Maxim Sølund Kirsebom, and Parthanil Roy. Continued fractions, the Chen-Stein method and extreme value theory. *Ergodic Theory Dyn. Syst.*, 41(2):461–470, 2021.

[10] Masaki Hirata, Benoît Saussol, and Sandro Vaienti. Statistics of return times: A general framework and new applications. *Communications in Mathematical Physics*, 206:33–55, 1999.

[11] Marius Iosifescu. Doeblin and the metric theory of continued fractions: A functional theoretic solution to Gauss' 1812 problem. In *Doeblin and modern probability. Proceedings of the Doeblin conference '50 years after Doeblin: development in the theory of Markov chains, Markov processes, and sums of random variables' held November 2-7, 1991 at the university of Tübingen's Heinrich Fabri Institut, Blaubeuren, Germany*, pages 97–110. Providence, RI: American Mathematical Society, 1993.

[12] Marius Iosifescu and Cor Kraaikamp. *Metrical Theory of Continued Fractions*. Mathematics and Its Applications. Springer Netherlands, 2002.

[13] Olav Kallenberg. *Random measures, theory and applications*, volume 77 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2017.

[14] J. F. C. Kingman. *Poisson processes*, volume 3 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1993. Oxford Science Publications.

[15] Leonid Kontorovich. *Measure concentration of strongly mixing processes with applications*. ProQuest LLC, Ann Arbor, MI, 2007. Thesis (Ph.D.)–Carnegie Mellon University.

[16] Leonid (Aryeh) Kontorovich and Kavita Ramanan. Concentration inequalities for dependent random variables via the martingale method. *The Annals of Probability*, 36(6):2126 – 2158, 2008.

[17] Günter Last and Mathew Penrose. *Lectures on the Poisson process*, volume 7 of *Institute of Mathematical Statistics Textbooks*. Cambridge University Press, Cambridge, 2018.

[18] Michel Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[19] Benjamin Weiss. Poisson generic points, 23-27 November 2020. Jean-Morlet Chair conference on Diophantine Problems, Determinism and Randomness. Centre International de Rencontres Mathématiques. Audio-visual resource: doi:10.24350/CIRM.V.19690103.

[20] Roland Zweimüller. Hitting times and positions in rare events. *Annales Henri Lebesgue*, 5:1361–1415, 2022.

Nicolás Álvarez
ICC
Universidad de Buenos Aires & CONICET Argentina
nialvarez@dc.uba.ar

Verónica Becher
Departamento de Computación, Facultad de Ciencias Exactas y Naturales & ICC
Universidad de Buenos Aires & CONICET Argentina
vbecher@dc.uba.ar

Eda Cesaratto
Universidad Nacional de Gral. Sarmiento & CONICET Argentina
ecesaratto@campus.ungs.edu.ar

Martín Mereb
Departamento de Matemática, Facultad de Ciencias Exactas y Naturales & IMAS
Universidad de Buenos Aires & CONICET Argentina
mmereb@gmail.com

Yuval Peres
Beijing Institute of Mathematical Sciences and Applications (BIMSA)
yuval@yuvalperes.com

Benjamin Weiss
Institute of Mathematics, Hebrew University of Jerusalem
Jerusalem, Israel
benjamin.weiss1@mail.huji.ac.il